Volume 4, Issue 12, December 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Review of a Methodology for Direct and Indirect Discrimination Prevention in Data Mining

Shubhangi R. Khade, Y. B. Gurav
Padmashri Vasantdada Patil Institute of Technology
Maharashtra, India

Abstract: ***Data mining is A more and more necessary technology for extracting useful data hidden in giant collections of information. There are, however, negative social perceptions concerning data processing, among that potential privacy invasion and potential discrimination. The latter consists of below the belt treating individuals on the premise of their happiness to a selected cluster. Automatic knowledge collection and data processing techniques like classification rule mining have paved the thanks to creating automatic selections, like loan granting/denial, premium computation, etc. If the coaching knowledge sets square measure biased in what regards discriminatory (sensitive) attributes like gender, race, religion, etc., discriminatory selections might result. For this reason, antidiscrimination techniques together with discrimination discovery and interference are introduced in data processing. Discrimination is often either direct or indirect. Direct discrimination happens once selections square measure created supported sensitive attributes. Indirect discrimination happens once selections square measure created supported no sensitive attributes that square measure powerfully correlative with biased sensitive ones. During this paper, we have a tendency to tackle discrimination interference in data processing and propose new techniques applicable for direct or indirect discrimination interference severally or each at an equivalent time. we have a tendency to discuss a way to clean coaching knowledge sets and outsourced knowledge sets in such the way that direct and/or indirect discriminatory call rules square measure reborn to legitimate (nondiscriminatory) classification rules. we have a tendency to conjointly propose new metrics to gauge the utility of the planned approaches and that we compare these approaches. The experimental evaluations demonstrate that the planned techniques square measure effective at removing direct and/or indirect discrimination biases within the original knowledge set whereas protective knowledge quality***

Index Term: ***Antidiscrimination, data mining, direct and indirect discrimination prevention, rule protection, rule generalization, Privacy***

I. INTRODUCTION

In social science, it involves denying to members of 1 cluster opportunities that are out there to different teams. there's an inventory of antidiscrimination acts, that are laws designed to forestall discrimination on the idea of variety of attributes (e.g., race, religion, gender, position, disability, legal status, and age) in numerous settings (e.g., employment and coaching, access to public services, credit and insurance, etc.). As an example, the ecru Union implements the principle of equal treatment between men and girls within the access to and provider of products and services in [3] or in matters of employment and occupation in [4]. Though there are some laws against discrimination, all of them arc reactive, not proactive. Technology will add proactively to legislation by contributory discrimination discovery and bar techniques.

Discrimination is either direct or indirect (also known as systematic). Direct discrimination consists of rules or procedures that expressly mention minority or underprivileged teams supported sensitive discriminatory attributes associated with cluster membership. Indirect discrimination consists of rules or procedures that, whereas not expressly mentioning discriminatory attributes, on purpose or accidentally might generate discriminatory selections. Redlining by money establishments (refusing to grant mortgages or insurances in urban areas they think about as deteriorating) is AN prototypical example of indirect discrimination, though in no way the sole one. With a small abuse of language for the sake of compactness, into his paper indirect discrimination also will be noted as redlining and rules inflicting indirect discrimination are going to be known as redlining rules [12]. Indirect discrimination might happen thanks to the supply of some background (rules), as an example that an explicit postal code corresponds to a deteriorating space or a neighborhood with largely black population. The background could be accessible from in public out there knowledge (e.g., census knowledge) or could be obtained from the initial knowledge set itself thanks to the existence of nondiscriminatory attributes that are extremely correlate with the sensitive ones within the original data set our study. Indirect discrimination might happen thanks to the supply of some background (rules).

II. RELATED WORK

Preprocessing. Transform the source data in such a way that the discriminatory biases contained in the original data are removed so that no unfair decision rule can be mined from the transformed data and apply any of the standard data mining algorithms.

In processing. Change the data mining algorithms in such a way that the resulting models do not contain unfair decision rules. For example, an alternative approach to cleaning the discrimination from the original data set is proposed in [2] whereby the non-discriminatory constraint is embedded into a decision tree learner by changing its splitting criterion and pruning strategy through a novel leaf rrelabeling approach.

III. A PROPOSAL FOR DIRECT AND INDIRECT

A. The Approach

Our approach for direct and indirect discrimination prevention can be described in terms of two phases:

- Discrimination measurement. Direct and indirect-discriminatory rules and redlining rules. To this end, first, based on predetermined discriminatory items in DB, frequent classification rules in FR are divided in two groups: PD and PND rules . Second, direct discrimination is measured by identifying _-discriminatory rules among the PD rules using a direct discrimination measure (elite) and a discriminatory threshold (_). Third, indirect discrimination is measured by identifying redlining rules among the PND rules combined with background knowledge, using an indirect discriminatory measure (elb), and a discriminatory threshold (_). Let MR be the database of direct _-discriminatory rules obtained with the above process. In addition, let RR be the database of redlining rules and their respective indirect _-discriminatory rules obtained with the above process. Discrimination discovery includes identifying
- Data transformation. Transform the original data DB in such a way to remove direct and/or indirect discriminatory biases, with minimum impact on the data and on legitimate decision rules, so that no unfair decision rule can be mined from the transformed data. In the following sections, we present the data transformation methods that can be used for this purpose.

B. Data Transformation for Direct Discrimination

In order to convert each _-discriminatory rule into an_-protective rule, based on the direct discriminatory measure (i.e., Definition 2), we should enforce the following inequality for each _-discriminatory rule r0 : A,B→ C in MR, where A is a discriminatory item set: that if each _-discriminatory rule r0 : A,B -> C in the database of decision rules was an instance of at least one no redlining (legitimate) PND rule r : D,B -> C, the data set would be free of direct discrimination.

IV. BACKGROUND

A. Basic Definitions

- A data set is a collection of data objects (records) and their attributes. Let DB be the original data set. An item is an attribute along with its value, e.g., Race = black.
- .An item set, i.e., X, is a collection of one or more items, e.g., {Foreign worker =Yes, City = NYC}.
- A classification rule is an expression X → C, where C is a class item (a yes/no decision), and X is an item set containing no class item, e.g., {Foreign worker = Yes; City = NYC} →Hire = no. X is called the premise of the rule.
- The support of an item set, supp(X), is the fraction of records that contain the item set X. We say that a rule X →C is completely supported by a record if both X and C appear in the record.
- The confidence of a classification rule, conf(X→C), measures how often the class item C appears in records that contain X. Hence, if supp(X)→ 0 the

Con *f(X--->C) =supp(X, C)/supp(X)*

$$conf(X \rightarrow C) = \frac{supp(X,C)}{supp(X)}.$$

V. ALGORITHMS

We describe in this section our algorithms based on the direct and indirect discrimination preventions methods proposed in Sections 3.2, 3.3, and 3.4. There are some assumptions common to all algorithms in this section. First, we assume the class attribute in the original data set DB to be binary (e.g., denying or granting credit). Second, we consider classification rules with ergative decision (e.g., denying credit) to be in FR. Third, we assume the discriminatory item sets (i.e., A) and the non-discriminatory item sets (i.e., D) to be binary or no binary categorical.

Direct discrimination prevention algorithms, we start with direct rule protection. Algorithm 1 details Method 1 for DRP. For each direct _-discriminatory rule r0 in MR (Step 3), after finding the subset DBc (Step 5), records in DBc should be changed until the direct rule protection requirement (Step 10) is met for each respective rule (Steps 10-14).

Algorithm 1. DIRECT RULE PROTECTION (METHOD 1)

Inputs: $\mathcal{DB}$, $\mathcal{FR}$, $\mathcal{MR}$, α, DI_s
Output: $\mathcal{DB}'$ (transformed data set)
for each $r' : A, B \rightarrow C \in \mathcal{MR}$ **do**
  $\mathcal{FR} \leftarrow \mathcal{FR} - \{r'\}$
  $\mathcal{DB}_c \leftarrow$ All records completely supporting $\neg A$, $B \rightarrow \neg C$
  for each $db_c \in \mathcal{DB}_c$ **do**
    Compute $impact(db_c) = |\{r_a \in \mathcal{FR} | db_c$ supports the premise of $r_a\}|$
  end for
  Sort $\mathcal{DB}_c$ by ascending impact
  while $conf(r') \geq \alpha \cdot conf(B \rightarrow C)$ **do**
    Select first record in $\mathcal{DB}_c$
    Modify discriminatory item set of db_c from $\neg A$ to A in $\mathcal{DB}$
    Recompute $conf(r')$
  end while
end for
Output: $\mathcal{DB}' = \mathcal{DB}$

Algorithm 2. DIRECT RULE PROTECTION (METHOD 2)

Inputs: $\mathcal{DB}$, $\mathcal{FR}$, $\mathcal{MR}$, α, DI_s
Output: $\mathcal{DB}'$ (transformed data set)
for each $r' : A, B \rightarrow C \in \mathcal{MR}$ **do**
  Steps 4-9 Algorithm 1
  while $conf(B \rightarrow C) \leq \frac{conf(r')}{\alpha}$ **do**
    Select first record in $\mathcal{DB}_c$
    Modify the class item of db_c from $\neg C$ to C in $\mathcal{DB}$
    Recompute $conf(B \rightarrow C)$
  end while
end for
Output: $\mathcal{DB}' = \mathcal{DB}$

Algorithm 3. DIRECT RULE PROTECTION AND RULE GENERALIZATION

Inputs: $\mathcal{DB}$, $\mathcal{FR}$, $\mathcal{TR}$, $p \geq 0.8$, α, DI_s
Output: $\mathcal{DB}'$ (transformed data set)
for each $r' : A, B \rightarrow C \in \mathcal{TR}$ **do**
  $\mathcal{FR} \leftarrow \mathcal{FR} - \{r'\}$
  if $TR_{r'} = RG$ **then**
    // Rule Generalization
    $\mathcal{DB}_c \leftarrow$ All records completely supporting $A,\ B, \neg D \rightarrow C$
    Steps 6-9 Algorithm 1
    while $conf(r') > \frac{conf(r_b : D, B \rightarrow C)}{p}$ **do**
      Select first record in $\mathcal{DB}_c$
      Modify class item of db_c from C to $\neg C$ in $\mathcal{DB}$
      Recompute $conf(r')$
    end while
  end if
  if $TR_{r'} = DRP$ **then**
    // Direct Rule Protection
    Steps 5-14 Algorithm 1 or Steps 4-9 Algorithm 2
  end if
end for
Output: $\mathcal{DB}' = \mathcal{DB}$

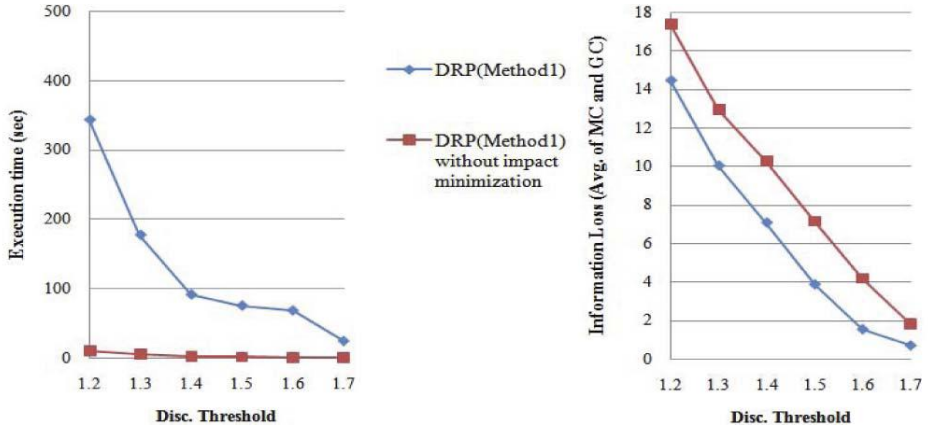
Direct and Indirect Discrimination Prevention Algorithms-

Algorithm 4. DIRECT AND INDIRECT DISCRIMINATION PREVENTION

Inputs: $\mathcal{DB}$, $\mathcal{FR}$, $\mathcal{RR}$, $\mathcal{MR}$, α, DI_s
Output: $\mathcal{DB}'$ (transformed data set)
for each $r : X \rightarrow C \in \mathcal{RR}$, where $D, B \subseteq X$ **do**
  $\gamma = conf(r)$
  for each $r' : (A \subseteq DI_s), (B \subseteq X) \rightarrow C \in \mathcal{RR}$ **do**
    $\beta_2 = conf(r_{b2} : X \rightarrow A)$
    $\Delta_1 = supp(r_{b2} : X \rightarrow A)$
    $\delta = conf(B \rightarrow C)$
    $\Delta_2 = supp(B \rightarrow A)$
    $\beta_1 = \frac{\Delta_1}{\Delta_2}$ $//conf(r_{b1} : A, B \rightarrow D)$
    Find $\mathcal{DB}_c$: all records in $\mathcal{DB}$ that completely support $\neg A, B, \neg D \rightarrow \neg C$
    Steps 6-9 Algorithm 1
    if $r' \in \mathcal{MR}$ **then**
      while $(\delta \leq \frac{\beta_1(\beta_2+\gamma-1)}{\beta_2 \cdot \alpha})$ and $(\delta \leq \frac{conf(r')}{\alpha})$ **do**
        Select first record db_c in $\mathcal{DB}_c$
        Modify the class item of db_c from $\neg C$ to C in $\mathcal{DB}$
        Recompute $\delta = conf(B \rightarrow C)$
      end while
    else
      while $\delta \leq \frac{\beta_1(\beta_2+\gamma-1)}{\beta_2 \cdot \alpha}$ **do**
        Steps 15-17 Algorithm 4
      end while
    end if
  end for
end for
for each $r' : (A, B \rightarrow C) \in \mathcal{MR} \setminus \mathcal{RR}$ **do**
  $\delta = conf(B \rightarrow C)$
  Find $\mathcal{DB}_c$: all records in $\mathcal{DB}$ that completely support $\neg A, B \rightarrow \neg C$
  Step 12
  while $(\delta \leq \frac{conf(r')}{\alpha})$ **do**
    Steps 15-17 Algorithm 4
  end while
end for
Output: $\mathcal{DB}' = \mathcal{DB}$

VI. CONCLUSION AND FUTURE WORK

It is more than obvious that most people do not want to be discriminated because of their gender, religion, nationality, age, and so on, especially when those attributes are used for making decisions about them like giving them a job, loan, insurance, etc. The purpose of this paper was to develop a new pre-processing discrimination prevention methodology including different data transformation methods that can prevent direct discrimination, indirect discrimination or both of them at the same time.



REFERENCES

[1] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," Proc. 20th Int'l Conf. Very Large Data Bases, pp. 487-499, 1994.

[2] T. Calders and S. Verwer, "Three Naive Bayes Approaches for Discrimination-Free Classification," Data Mining and Knowledge Discovery, vol. 21, no. 2, pp. 277-292, 2010.

[3] European Commission, "EU Directive 2004/113/EC on Anti- Discrimination," http://eur-lex.europa.eu/LexUriServ/ LexUriServ.do?uri=OJ:L:2004:373:0037:0043:EN:PDF, 2004.

[4] European Commission, "EU Directive 2006/54/EC on Anti- Discrimination," http://eur lex.europa.eu/LexUriServ/ LexUriServ.do?uri=OJ:L:2006:204:0023:0036:en:PDF, 2006.

[5] S. Hajian, J. Domingo-Ferrer, and A. Martı´nez-Balleste´, "Discrimination Prevention in Data Mining for Intrusion and Crime Detection," Proc. IEEE Symp. Computational Intelligence in Cyber Security (CICS '11), pp. 47-54, 2011.

[6] S. Hajian, J. Domingo-Ferrer, and A. Martı´nez-Balleste´, "Rule Protection for Indirect Discrimination Prevention in Data Mining," Proc. Eighth Int'l Conf. Modeling Decisions for Artificial Intelligence (MDAI '11), pp. 211-222, 2011.

[7] F. Kamiran and T. Calders, "Classification without Discrimination," Proc. IEEE Second Int'l Conf. Computer, Control and Comm. (IC4 '09), 2009.

[8] F. Kamiran and T. Calders, "Classification with no Discrimination by Preferential Sampling," Proc. 19th Machine Learning Conf. Belgium and The Netherlands, 2010.

[9] F. Kamiran, T. Calders, and M. Pechenizkiy, "Discrimination Aware Decision Tree Learning," Proc. IEEE Int'l Conf. Data Mining (ICDM '10), pp. 869-874, 2010.

[10] R. Kohavi and B. Becker, "UCI Repository of Machine Learning Databases," http://archive.ics.uci.edu/ml/datasets/Adult, 1996.

[11] D.J. Newman, S. Hettich, C.L. Blake, and C.J. Merz, "UCI Repository of Machine Learning Databases," http://archive. ics.uci.edu/ml, 1998.

[12] D. Pedreschi, S. Ruggieri, and F. Turini, "Discrimination-Aware Data Mining," Proc. 14th ACM Int'l Conf. Knowledge Discovery and Data Mining (KDD '08), pp. 560-568, 2008.

[13] D. Pedreschi, S. Ruggieri, and F. Turini, "Measuring Discrimination in Socially-Sensitive Decision Records," Proc. Ninth SIAM Data Mining Conf. (SDM '09), pp. 581-592, 2009.

[14] D. Pedreschi, S. Ruggieri, and F. Turini, "Integrating Induction and Deduction for Finding Evidence of Discrimination," Proc. 12th ACM Int'l Conf. Artificial Intelligence and Law (ICAIL '09), pp. 157- 166, 2009.

[15] S. Ruggieri, D. Pedreschi, and F. Turini, "Data Mining for Discrimination Discovery," ACM Trans. Knowledge Discovery from Data, vol. 4, no. 2, article 9, 2010.

[16] S. Ruggieri, D. Pedreschi, and F. Turini, "DCUBE: Discrimination Discovery in Databases," Proc. ACM Int'l Conf. Management of Data (SIGMOD '10), pp. 1127-1130, 2010.

[17] P.N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Addison-Wesley, 2006.

[18] United States Congress, US Equal Pay Act, http://archive. eeoc.gov/epa/anniversary/epa-40.html, 1963.

[19] V. Verykios and A. Gkoulalas-Divanis, "A Survey of Association Rule Hiding Methods for Privacy," Privacy-Preserving Data Mining: Models and Algorithms, C.C. Aggarwal and P.S. Yu, eds., Springer, 2008.

